JOSU MEZO

Encuestas: el margen de error, la paja y el grano

En el primer artículo de esta sección (número 6 de *Cuadernos de Periodistas)* veíamos cómo, contra lo que la intuición nos sugiere, es posible encuestar a sólo unos cientos de entrevistados, o a un millar, para averiguar la opinión de centenares de miles o millones de personas. Como veíamos, la cuestión clave es la aleatoriedad, de manera que una buena práctica periodística debería rechazar las encuestas con muestra no aleatoria como temas centrales de una noticia. Si consiguiéramos eso, ya habríamos eliminado el error más grosero que cometen los medios en relación con las encuestas.

Sin embargo, hay un segundo grupo de errores, también muy comunes, que afectan a las encuestas *buenas,* las que se basan en muestras aleatorias, y que tiene que ver con la mala comprensión y manejo de un concepto del que les hablaba ya en el artículo anterior: el margen de error. En efecto, cuando decimos que una muestra aleatoria permite averiguar, preguntando a unas pocas personas, lo que piensan millones, lo hacemos con una cautela importantísima: el resultado de la encuesta no es idéntico al que obtendríamos si pudiéramos entrevistar a toda la población, sino aproximado. Precisamente el margen de error es lo que nos sirve para calcular, a partir del dato que obtenemos en la muestra, cuál debe de ser el dato real en la población.

Podemos explicar el margen del error, entonces, como un número que tenemos que sumar y restar al resultado obtenido en una muestra, para estimar el rango u horquilla de datos dentro del cual se encontrará el verdadero dato en la población. Así, si en una encuesta obtenemos que un 32% de los entrevistados se declara consumidor de un determinado producto, y el margen de error es de ±4%, con un 95% de confianza, lo que esto quiere decir es que podemos afirmar, con un 95% de probabilidades de acertar, que si pudiéramos entrevistar a toda la población, los que consumen ese producto estarían entre el 28% (32−4) y el 36% (32+4).

¿Y cómo se sabe el margen de error? De momento, nos quedaremos con el que se suele publicar en la ficha técnica de los estudios, con un texto como este: "el margen de error es ±3,5% para *p*=*q*=0,5 y con un nivel de confianza del 95%". Como vimos en el artículo anterior, "*p*=*q*=0,5" se refiere a la proporción de entrevistados que escoge una respuesta *(p)* y a los que no la escogen *(q)*. Cuando esta es 0,5 (50%) es cuando el margen de error es mayor, y por ello ése es el que se incluye en la ficha técnica (es preferible que se sobreestime el margen de error a que se infra estime). Cuando el porcentaje de entrevistados que elige una respuesta es menor o mayor del 50%, el margen de error disminuye, pero lo hace muy suavemente, así que en la mayor parte de los casos, como lectores, nos podemos conformar con el margen de error publicado en la ficha técnica (más abajo presentaré una excepción a esta regla)*.

Lógicamente, cuando se realiza una encuesta no nos interesa saber lo que piensan o hacen las 1.000 ó 2.000 personas entrevistadas, sino el conjunto de la población. Por ello, estrictamente hablando, el resultado de las encuestas debería expresarse en forma de rangos. Si una encuesta tiene un margen de error de ±3% (con un nivel de confianza del 95,5%), en lugar de decir, al resumirla, que el partido A obtiene un 38% de los apoyos habría que decir que entre el 35% y el 41% de la población apoya al partido A (con un 95,5% de probabilidad).

¿Por qué no lo hacemos así? Hay al menos dos razones. La primera es de economía del lenguaje. Sería una pesadez dar cada dato con una horquilla, así que, en su lugar (cuando las cosas se hacen bien), se adjunta al estudio una ficha técnica que informa del margen de error, y se hace la suposición bienintencionada (y errónea, como veremos luego) de que los lectores y usuarios de la encuesta sabrán hacer el cálculo mental sobre la horquilla. Hay también una razón matemática: no todos los datos de la horquilla tienen las mismas probabilidades de ser ciertos. No es posible entrar aquí en detalles técnicos sobre esto, pero las mismas reglas de la probabilidad que nos permiten calcular que hay un 95,5% de probabilidades

En las encuestas, no todos los datos de la horquilla tienen las mismas probabilidades de ser ciertos.

de que el porcentaje real en la población esté entre el 35% y el 41% nos permiten también calcular la probabilidad de cualquier otro intervalo de valores, generando una tabla como la siguiente.

Probabilidades de diferentes rangos de valores en la población, para un valor del 38% obtenido en una muestra de 1.047 entrevistas

RANGO DE VALORES	PROBABILIDAD
Hasta 34%	0,38%
34% a 35%	1,89%
35% a 36%	6,85%
36% a 37%	16,13%
37% a 38%	24,75%
38% a 39%	24,75%
39% a 40%	16,13%
40% a 41%	6,85%
41% a 42%	1,89%
Más de 42%	0,38%

Como puede verse, los tramos de valores más probables son los 'centrales', en torno al valor obtenido en la muestra. Así, los dos puntos centrales (de 37% a 39%) tienen una probabilidad acumulada de casi el 50%, si ampliamos el rango un punto en cada dirección (desde 36% a 40%), la probabilidad llega hasta casi el 82%, y si extendemos el rango hasta los seis puntos (desde el 35% al 41%), la probabilidad crece hasta el 95,5% (por eso decimos que, para un nivel de confianza del 95,5%, el margen de error es ±3%). Con un rango de ocho puntos (desde 34% a 42%), la probabilidad sería ya de más del 99%.

La decisión de optar por el 95,5% de confianza, o por el 99%, o por el 80% es relativamente arbitraria. Todo depende de cuán seguro quiera estar uno de sus conclusiones. Lo más convencional en ciencias sociales es aceptar el 95% (o a veces el 95,5%) de confianza, que se puede entender, al revés, como un máximo de un 5% de probabilidades de equivocarse (o una entre veinte). Ese es el nivel de confianza que suele aparecer también en la ficha técnica de las encuestas que encargan y publican los propios medios de comunicación (y el que usaré también en este artículo, cuando no se diga otra cosa).

El problema de las comparaciones

En todo caso, el valor más probable, individualmente considerado, sería precisamente el que se ha obtenido en la muestra, y por eso no es del todo incorrecto utilizarlo cuando resumimos los resultados. ¿Quizá entonces no es tan grave usar simplemente el dato de la muestra cuando se resume una encuesta, o se hace un titular? No lo sería, si no fuera porque los datos de las encuestas no se utilizan casi nunca de uno en uno, sino comparativamente. Por ejemplo, las encuestas electorales suelen informar sobre cuántos puntos saca el partido A al partido B, cuánto ha subido o bajado el apoyo al partido A desde la encuesta anterior, o en qué comunidades autónomas el apoyo al partido A es mayor o menor. O si la encuesta es sobre un asunto de actualidad (por ejemplo, una propuesta política para aprobar una ley), es también muy común que se informe de si son más los que están a favor o en contra, si los favorables a la medida en cuestión son más o menos que en estudios anteriores, o cuáles son los porcentajes en distintas partes del país, o entre los votantes de diferentes partidos.

¿Y cómo se suelen hacer esas comparaciones? Pues normalmente se realizan utilizando exclusivamente los datos precisos obtenidos de la muestra, sin tener en cuenta los márgenes de error. Y aquí es cuando este olvido puede llevar a sacar conclusiones erróneas, o cuando menos, arriesgadas (con una probabilidad no despreciable de equivocarse).

Para entender esto, pensemos en una situación en la que sabemos de antemano el porcentaje en la población que tiene una determinada característica, por ejemplo, que un 52% son mujeres. Si realizamos una encuesta con un margen de error de ±4%, tenemos un 95% de probabilidades de que en la muestra no haya menos de un 48% de mujeres ni más de un 56%. De aquí se deduce una consecuencia lógica que muchas veces se olvida: dos encuestas hechas el mismo día en el mismo lugar y por el mismo procedimiento podrían encontrar un 49% y un 53% de mujeres, y no habría en ello nada particular. Entra dentro del margen de error.

Lo mismo sucedería con cualquier otra característica de la que no conozcamos su proporción real en la población. Imaginemos que dos periódicos publican el mismo día dos encuestas electorales, hechas por un exquisito procedimiento de muestra aleatoria, con 1.000 entrevistados (margen de error de ±3,1%) y una otorga a un partido un 40% de intención de voto y otra un 38% (suponemos que en ambos casos se tratara de respuestas directas de los entrevistados). No habría en ello nada de sorprendente ni anómalo, ya que ambos resultados son perfectamente compatibles con un rango de valores reales en la población bastante amplio (aproximadamente entre 37% y 41%).

Bien, pero ahora pensemos que casi todos nuestros principales periódicos publican con cierta regularidad encuestas propias (es decir, encargadas y pagadas por el propio periódico) orientadas sobre todo a temas políticos y en las que no suele faltar la estimación del posible voto de los ciudadanos en la siguiente consulta electoral, ya sea nacional o autonómica. Los resultados de estas encuestas se suelen publicar en domingo, el día de mayor difusión, y ocupan muchas veces posiciones privilegiadas en la portada, y amplios espacios en las páginas de actualidad nacional.

Es muy habitual que estas encuestas se realicen con 1.000 entrevistados, lo que daría un margen de error (cuando p=q=50%) de 3,16%. Ahora imaginemos que, con dos meses de diferencia, el mismo periódico o cadena de radio hace dos encuestas similares, con 1.000 entrevistados, y en la primera obtiene que el apoyo al partido A es del 38% y en la segunda obtiene que el apoyo es del 40%. ¿Habría algún periódico o periodista que se resistiera a decir que el apoyo electoral al partido A, según su encuesta, había subido un 2%? Seguramente ninguno (todo hay que decirlo: tampoco muchos sociólogos se resistirían). Pero dado el margen de error, no podemos saber con seguridad ni siquiera si la intención de voto al partido A ha subido. Podría ser que la intención de voto fuera estable, y que simplemente por azar nos hubieran salido dos muestras con esa diferencia.

Y sin embargo, es muy común que estos cambios inciertos en la intención de voto se presenten como *hechos reales,* e incluso se intenten explicar con reflexiones sobre los acontecimientos políticos recientes. Al partido A le ha perjudicado (o beneficiado) tal o cual estrategia, o tal o cual decisión, el debate en el parlamento sobre esto, el escándalo sobre aquello. En realidad, se está poniendo un gran esfuerzo en explicar algo que no sabemos si existe, porque, con muestras de 1.000 entrevistados, una subida o bajada de uno, dos o tres puntos, puede deberse perfectamente al azar.

> Es muy común que cambios inciertos en la intención de voto se presenten como *hechos reales.*

En términos técnicos se diría que la diferencia entre las dos encuestas no es *estadísticamente significativa,* es decir, que no es lo suficientemente grande para estar razonablemente seguros de que responde a un fenómeno real en la población, y no al azar. Dicho de otra forma: no tenemos un 95% de confianza (o un 99%, o el porcentaje convencionalmente escogido) de que la intención de voto al partido A sea en la población mayor en el momento de la segunda encuesta que en el momento de la primera encuesta. O alternativamente: hay una probabilidad mayor del 5% de que la diferencia entre las dos encuestas se deba al azar.

Quizá alguien pueda pensar que esto no es tan grave. Si todas las cosas que se cuentan en los medios fueran seguras con un 94% de probabilidad, no estaríamos tan mal. Quizá no deba exigirse para hacer una información de prensa una seguridad del 95% y baste con un 90%, o un 80%. En efecto, puede argumentarse que para la 'verdad periodística' podríamos aceptar un nivel de confianza menos estricto que para la 'verdad científica'. Pero en la práctica lo que sucede, me temo, es que la mayoría de los periodistas desconocen estas cuestiones y simplemente escriben, o hablan, como si el margen de error no existiese, informando con la misma relevancia y atribuyéndole la misma verosimilitud a diferencias con un nivel de confianza del 99%, el 60% o el 30%.

Ejemplos poco edificantes

Vamos a ver unos cuantos casos reales de *noticias* en las que se han presentado a los lectores o espectadores como datos ciertos resultados de encuestas sin tener en cuenta la *significación estadística,* es decir, sin tener en cuenta que había probabilidades no desdeñables (en algunos casos enormes) de que la diferencia entre dos datos provenientes de encuesta se debiera en realidad al azar, y no tuviera ningún soporte en la realidad.

Ya hemos visto que esto sucede en muchos casos cuando se comparan dos encuestas sucesivas sobre intención de voto. Para salirnos de lo político podemos utilizar un ejemplo muy divertido del *Abc* de mayo de 2005. El titular decía: 'Crece en España por primera vez desde 2001 el número de personas declaradas católicas'. La noticia afirmaba que tras una constante caída desde 2001 de las personas que se declaraban católicas, las encuestas del CIS reflejaban que en el último año esa tendencia se había frenado e incluso invertido, aunque muy levemente. La evidencia era la siguiente: los que se declararon católicos fueron el 79,1% en junio de 2004 y el 79,3% en marzo de 2005.

Pero este ligerísimo aumento olvida por completo el margen de error, y el intervalo de confianza. En realidad, teniendo en cuenta que ambas encuestas tenían 2.500 entrevistados, el margen de error para estos datos sería ±1,6%, de forma que el resultado de junio de 2004 debería ser leído como un intervalo entre 77,5% y 80,7% (79,1% ±1,6%) y el dato de 2005 como un intervalo entre 77,7% y 80,9% (79,3% ±1,6%). Los intervalos de confianza no es ya que se superpongan parcialmente, sino que son prácticamente idénticos en todo su recorrido. Con ese resultado, es simple y llanamente imposible decir que el número de españoles que se declararían católicos si los pudiéramos entrevistar a todos sería más alto en 2005 que en 2004. Tampoco lo contrario, por supuesto. La diferencia, evidentemente, no es estadísticamente significativa. Para que una diferencia tan pequeña fuera significativa, de hecho, haría falta una muestra gigantesca (algo así como 300.000 entrevistados). En definitiva, una variación de 0,2% entre dos encuestas con un margen de error de ±1,6%, debe interpretarse como: no hay cambios, o si los hay son tan pequeños que la encuesta no los puede detectar.

No es necesario siquiera que estemos ante dos encuestas distintas para que se nos presente este problema. También sucede cuando se comparan, en una misma encuesta, los porcentajes correspondientes a dos respuestas de una misma pregunta: los que están a favor o en contra de una determinada política, los que tienen opinión favorable o desfavorable de un personaje, o los que van a votar a un partido u a otro. Por ejemplo, el año pasado, en pleno debate sobre el Plan Ibarretxe la cadena Ser publicó en su página web los resultados de una encuesta hecha en el País Vasco, con el siguiente titular: "El Plan Ibarretxe, rechazado en el País Vasco por un escaso margen". Los resultados eran estos, según el titular de *El País:* 'El 42% de los vascos rechaza el Plan Ibarretxe frente al 39% que lo apoya'. ¿No era correcto, con esos datos, el titular de la Ser? No, no lo era, ni tampoco realmente el de *El País,* pues ambos olvidaban el margen de error.

La encuesta se basaba en una muestra de 400 personas, lo que daría un margen de error de ±4,9%. Por tanto, todo lo que se podía decir, con un 95,5% de probabilidad, era que los que se oponían al Plan Ibarretxe en el País Vasco estaban entre el 37,4% y el 47,2% (42% ±4,9%); y los que lo apoyaban serían entre el 34,1% y el 42,9% (39% ±4,9%). Ambas horquillas tienen un recorrido de 9,8 puntos (4,9 x 2) y comparten una amplio tramo en común. Por lo tanto, con esos datos no se podría afirmar con demasiada confianza si en la población del País Vasco había más gente a favor o en contra del Plan Ibarretxe.

¿Cuál sería la probabilidad de que en realidad los favorables al plan fuera más que los contrarios? Mi estimación es que esa probabilidad no estaría lejos del 25%. Es decir, el titular tenía aproximadamente un 25% de probabilidades de estar completamente equivocado, y un 75% de probabilidades de acertar, al menos, sobre qué grupo era mayor. No es catastrófico, pero es suficientemente malo para evitarlo. Hubiera sido mucho mejor titular "Los vascos, muy divididos ante el Plan Ibarretxe" y luego explicar en el texto de la noticia que en la encuesta aparecían más contrarios que favorables, pero la diferencia era pequeña y por tanto, debido al margen de error, aunque era probable que hubiera más gente en contra que a favor, no era algo seguro (todo esto, claro, desde un punto de vista simplemente matemático, y sin entrar en los problemas que puedan derivar de que las respuestas no sean sinceras).

Otra comparación errónea por ol-

¿Cree que Catalunya [sic] es solidaria o insolidaria con el resto de las comunidades autónomas? (Porcentajes verticales). Según comunidad autónoma de residencia

	TOTAL	CATAL.	ANDAL.	CASTILLAS	C. VAL.	GALICIA	MADRID	P. VASCO	RESTO
Solidaria	38,4	80,4	20,9	19,2	30,1	32,4	25,4	64,2	33,2
Insolidaria	49,4	14,1	64,5	64,4	59,2	54,4	61,2	22,3	53,4
Ns/Nc	12,2	4,4	14,6	16,4	10,7	13,2	13,4	13,2	13,5
(N)	1.000	160	171	101	106	65	136	50	211
Margen de error	±3,16	±7,91	±7,65	±9,94	±9,70	±12,39	±8,59	±14,14	±6,89

vidar el margen de error es la que se produce cuando dentro de una encuesta se comparan los porcentajes de respuestas a una misma pregunta entre diferentes subgrupos de la población. Por ejemplo, es muy frecuente que los resúmenes de encuestas comparen las proporciones de los que escogen una determinada respuesta entre los hombres y las mujeres, entre los jóvenes, adultos o más ancianos, entre los de diferentes comunidades autónomas, o entre los votantes de diferentes partidos. La situación se parece a la comparación entre dos encuestas, pero con una peculiaridad: al tratarse de submuestras, el margen de error ha de ser recalculado, teniendo en cuenta cuántos entrevistados componen cada una de las submuestras.

Un ejemplo, entre muchos, lo proporciona una encuesta publicada en febrero de 2006, en *El Periódico de Catalunya,* que comparaba las respuestas a la pregunta "¿Cree que Cataluña es solidaria o insolidaria con el resto de las comunidades autónomas?" de los encuestados catalanes, andaluces, madrileños, vascos, valencianos, gallegos, castellanos (de las dos Castillas) y otros. La tabla publicada era como la que figura en cabeza de la página (excepto las dos últimas líneas, que he añadido yo).

El comentario del periódico decía lo siguiente: "Son los andaluces (el 64,5%) los que en mayor medida acusan a los catalanes de insolidarios, seguidos muy de cerca por los encuestados de ambas Castillas (el 64,4%) y los madrileños (el 61,2%). Por detrás, pero aún con mayoría absoluta, están los valencianos (el 59,2%), los gallegos (el 54,4%) y los del resto de España (el 53,4%), con excepción de los propios catalanes (14,1%) y los vascos (22,6%)."

Una vez más, el margen de error es ignorado. Como se trataba de una encuesta con 1.000 entrevistados, que según la ficha técnica tenía un margen de error de ±3,16%, alguien podría pensar que para calcular las horquillas de datos habría que sumar y restar a cada porcentaje un 3,16%. Pero se equivocaría. Porque cada uno de esos porcentajes se refiere a una submuestra, de forma que habría que calcular por separado el margen de error para cada una de las comunidades autónomas. La tabla original no informaba de cuántos entrevistados hubo en cada comunidad, pero la ficha técnica indicaba que la muestra se hizo estratificada por comunidades autónomas, con lo que he podido hacer una estimación aproximada del número de entrevistas en cada comunidad, y del margen de error, que aparecen en las dos últimas filas.

Si se suma y resta a cada respuesta su margen de error, todas las horquillas de datos se pisarían (algunas son casi idénticas), excepto las correspondientes al País Vasco y Cataluña. Por tanto, con esa encuesta sólo puede decirse que los catalanes piensan por abrumadora mayoría que Cataluña es solidaria, que también lo piensan una mayoría sólida entre los vascos, y que en las demás partes de España hay una mayoría igualmente clara que piensa que Cataluña es insolidaria.

Tanto los periódicos como las empresas de opinión pública saben que con encuestas de 1.000 entrevistas casi todas las oscilaciones entre encuestas no son significativas.

Errores similares a este los cometen todos y cada uno de nuestros periódicos cuando analizan la opinión ante cualquier asunto político, con muestras totales de 1.000 entrevistados, y distinguen entre los votantes de PSOE, PP, IU, y si se descuida uno, CiU, o PNV. Dado que el voto a IU en las últimas elecciones fue aproximadamente un 3,7% del censo, y suponiendo que nadie ocultara su voto, la submuestra de votantes de IU oscilaría en torno a los 40 entrevistados, lo que tendría un margen de error de ±16%. Mi ejemplo favorito, en este sentido, lo dio *El Mundo,* hace ya unos años, cuando, tras el Pacto de Progreso en Baleares, que permitió gobernar al PSOE con el apoyo de varios pequeños partidos, hizo un titular de portada con una encuesta según la cual el 30% de los votantes de Unión Mallorquina desaprobaban el acuerdo (frente a un 47% a favor). La noticia no informaba del número de encuestados, pero aún suponiendo que fueran 1.000 (muchos para una encuesta de urgencia, y sólo de ámbito regional), los votantes de Unión Mallorquina en la encuesta deberían rondar los 40, y el margen de error por tanto estaría en torno a ±15%. Podría ser incluso que una mayoría estuviese en contra del pacto, pero con una encuesta general a toda la población sería imposible saberlo.

En busca de soluciones

¿Cómo se pueden evitar todos estos errores? La respuesta no es sencilla porque calcular qué diferencias entre valores de encuesta son significativas y cuáles no puede ser una tarea bien compleja, que requiere incluso, según el tipo de muestreo, conocimientos estadísticos avanzados. Así que no se puede esperar de un periodista con formación general que haga ese cálculo.

Sin embargo, las empresas periodísticas, como clientes que encargan muchos estudios de opinión pública, sí podrían fácilmente exigir a las empresas que los realizan que les informen adecuadamente sobre la significación de los resultados. Esas empresas tienen en sus filas expertos con formación estadística suficiente para poder averiguar qué diferencias entre respuestas, dentro de la encuesta, y que diferencias entre una encuesta y las anteriores son lo suficientemente grandes como para poder construir en torno a ellas una noticia. Esto no quiere decir que no se publique información alguna que no sea significativa al 95%. Como he dicho antes, puede ser aceptable, a efectos periodísticos, un grado de certidumbre menor, con dos matizaciones: que se fije un límite por debajo del cual no se publicarán resultados (¿un 90%? ¿un 80%?), y que se advierta al lector cuando el nivel de significación sea bajo.

¿Por qué no sucede esto ahora? Quizá el problema sea que tanto los periódicos como las empresas de opinión pública sepan que con encuestas de 1.000 entrevistas, como las que suelen hacer a menudo (que tienen un margen de error de aproximadamente ±3,2%), casi todas las oscilaciones entre encuestas, por ejemplo en la intención de voto, resultarían no significativas (porque raramente la intención de voto cambia en poco tiempo más allá de uno, dos o tres puntos). Si esto se reconociera, los medios no podrían llenar tanto espacio y tiempo con estas encuestas, ni las empresas facturar tanto por hacerlas. Así que quizá ambos estén interesados en mantener el *statu quo.*

Pero los medios publican también, por supuesto, resultados de muchas encuestas ajenas. ¿Qué hacer entonces? En primer lugar, siempre que fuera posible (cuando la fuente no es un distante instituto de otro país, sino una empresa o entidad cercana, que incluso convoca una rueda de prensa para contar los resultados), los periodistas podrían pedir a la fuente emisora información sobre la significación estadística de la información. A un investigador serio esta demanda no le puede resultar esotérica, ni habría de tener mayor problema en responder a ella. Si esta demanda fuera repetida y firme, por parte de los periodistas, se podría ir creando el uso, entre todos los productores habituales de información de encuestas (INE, CIS, institutos de opinión autonómicos, universidades, estudios de audiencia...), de ofrecer rutinariamente la significación estadística de las diferencias entre valores, así como los márgenes de error para las submuestras.

La comparación de submuestras requiere calcular previamente los márgenes de error.

Pero incluso si no es posible acceder a la fuente, ni pedirle explicaciones, se pueden dar algunas reglas sencillas que evitarán los errores más gruesos. Para empezar, podemos decir lo siguiente: si dos valores están muy distantes entre sí, más de dos veces el margen de error, entonces podemos estar seguros de que la diferencia es significativa, ya que las horquillas de datos resultantes "no se pisan". Al contrario, si la diferencia entre dos valores es menor o igual que el margen de error, es casi seguro que la diferencia no sea estadísticamente significativa, ya que las horquillas de datos se superponen en una buena parte.

La zona gris se produce cuando la diferencia entre los valores está entre una y dos veces el margen de error, ya que, aunque dos horquillas de valores se pisen, si lo hacen 'por poco', es posible que la diferencia sí sea significativa. Esto se debe a que, como hemos visto más arriba, no todos los valores de la horquilla son igualmente probables. Los valores extremos son relativamente poco probables, de forma que si dos rangos de datos se superponen sólo por sus extremos, es posible que la diferencia entre los dos valores sí sea estadísticamente significativa (es decir, la probabilidad de que responda a una diferencia real en la población sea mayor del 95%). Como no hay una fórmula matemática sencilla para resolver esta duda, el único consejo que se puede dar es el de la prudencia: si son datos sobre algo muy importante, infórmese de ello, pero siempre indicando al lector que no se está muy seguro de la significación estadística.

Para esas operaciones podemos basarnos en el margen de error incluido en la ficha técnica. Esto implica una ligera sobre-estimación del error cuando el porcentaje es distinto al 50%, pero para esos cálculos aproximados es suficiente. Ahora bien, si los porcentajes de los que hablamos se refieren a submuestras (porcentajes de los votantes de diferentes partidos, de distintas zonas de España, de hombres y mujeres, de jóvenes y viejos...), el margen de error sería mucho mayor, y no suele aparecer nunca en la ficha técnica.

En esos casos, si no queremos cometer errores de calado, tendremos que hacer un cálculo propio, aunque sea muy grueso, del margen de error. No se asusten, que es fácil. Lo único que necesitamos es tener una idea de cuál puede ser el tamaño de la submuestra, que podemos sacar de las propias respuestas a la encuesta, o podemos estimar aproximadamente. Por ejemplo, si las submuestras se refieren a diferentes partes de España, asumiendo que la muestra se ha hecho proporcionalmente en todo el territorio (es lo habitual), podemos calcular el tamaño de las submuestras aplicando a la muestra total el porcentaje de la población española que representan los distintos territorios. Si se trata de votantes de distintos partidos, podemos aplicar a la muestra el porcentaje del censo que votó a ese partido en las últimas elecciones. Calculado el tamaño aproximado de la submuestra, que llamaremos n, la fórmula para el margen de error, con un 95,5% de confianza, cuando $p=q=0,5$, es muy sencilla: es la raíz cuadrada de $(10.000/n)$. El resultado será impreciso, pero siempre mucho mejor que suponer que el margen de error es el mismo que el del total de la muestra.

Esa misma fórmula, por cierto, la podemos usar cuando no tenemos la ficha técnica del estudio. Esto no debería pasar nunca en una información de primera mano de una encuesta bien hecha (con una muestra representativa), por una institución competente y profesional. Pero a veces sucede que la información llega en un pequeño resumen, de segunda o tercera mano, en una nota de agencia, o en una noticia de otro medio. En ese caso, tras cerciorarnos de que realmente estamos ante una encuesta *buena,* y suponiendo que sepamos el tamaño de la muestra, podemos aplicar la fórmula anterior.

Así, imaginemos que nos llega información de una encuesta a 500 personas, según la cual el 41% de los entrevistados están a favor y el 35% en contra de una determinada propuesta política (el resto no sabe o no contesta). No tenemos la ficha técnica, aunque sí parece que se trata de una encuesta con muestra representativa. El margen de error aproximado sería la raíz de [10.000/500], es decir, 4,47%. La diferencia entre los dos valores está entre una y dos veces el margen de error (más exactamente es 1,34 veces el margen de error), por lo que está en la zona gris: si el tema es importante podemos informar sobre esa diferencia, pero aconsejando al lector que la tome con precaución, ya que pudiera ser producto del azar.

Recapitulación: cinco mensajes básicos

1. Por comodidad resumimos los datos de encuesta como si fueran datos precisos, pero en realidad, los resultados son rangos: podemos decir con un nivel de confianza determinado (normalmente el 95%) que, si pudiéramos entrevistar a toda la población, el porcentaje de personas que daría una determinada respuesta, estaría en el rango delimitado por el resultado de la encuesta más/menos el margen de error.

2. Cuando comparamos dos datos de encuesta (de la misma o de varias), por tanto, no comparamos realmente datos precisos, sino aproximaciones.

3. Las empresas e instituciones que hacen encuestas saben calcular cuándo la diferencia entre dos datos de encuesta es estadísticamente significativa, es decir, cuando la diferencia es suficiente para afirmar, con el 95% de confianza, que la diferencia no se debe al azar, y los datos son también distintos en la población. Siempre que sea posible, deberíamos pedir a los productores de encuestas que nos indicaran qué diferencias son significativas y cuáles no.

Una chuletilla de medio folio puede evitar bastantes disparates.

4. Si esto no es factible, la regla más simple es que no son significativas las diferencias entre valores menores que el margen de error, y sí son significativas las diferencias mayores que dos veces el margen de error. Las situaciones intermedias son dudosas.

5. El margen de error para los cálculos anteriores lo podemos tomar de la ficha técnica. Si no hay ficha técnica, o esta no informa del margen de error para submuestras, podemos calcular el margen de error aproximado como la raíz cuadrada de (10.000/*n*), siendo *n* el tamaño de la muestra o submuestra.

Ya ven, una chuletilla de medio folio podría evitar unos cuantos disparates que nos vemos obligador a leer y oír cotidianamente en nuestros medios. Por favor, pínchenla en el corcho más cercano.

*Si quisiéramos ser más puntillosos, hay calculadoras en Internet (en inglés, pero muy sencillas), para calcular el margen de error exacto, jugando con los tres datos básicos que son el número de miembros de la muestra *(sample size)*, el porcentaje obtenido en la muestra *(sample percentage)*, y el nivel de confianza *(confidence interval)*. Véase, por ejemplo, http://www.dssresearch.com/toolkit/secalc/error.asp